INRIA

Project-Team SIGNES

Linguistic signs, grammar and meaning: computational logic for natural language

Bordeaux - Sud-Ouest

THEME SYM

Activity Report

2008

Table of contents

SIGNES is an INRIA Project-Team joint with University of Bordeaux (UB1 and UB3 (TELEM, EA 4195)) and CNRS (CLLE, UMR 5263 and LaBRI, UMR 5800)

1. Team

Research Scientist
Gérard Huet [INRIA, Research Director (DR) located in Rocquencourt, HdR]
Renaud Marlet [INRIA, Research Associate (CR), HdR]
Richard Moot [CNRS, Research Associate (CR)]
Sylvain Salvati [INRIA, Research Associate (CR)]

Faculty Member
Lionel Clément [University Bordeaux 1, Assistant Professor (MCF)]
Anna Kupść [University Bordeaux 3, Assistant Professor (MCF)]
Christian Retoré [University Bordeaux 1, Team Leader, Professor (Pr), HdR]

External Collaborator
Maxime Amblard [PostDoctoral Fellow, France Télécom until Sep. 2008]
Christian Bassac [University Lyon 2, external collaborator since Sep. 2008, University Bordeaux 3 till Sep. 2008, Professor (Pr), HdR]
Mauro Gaio [University of Pau et des Pays de l'Adour, Professor (Pr), HdR]
Kim Gerdes [University Paris 3, Lecturer (MCF)]
Alain Lecomte [University Paris 8, Professor (Pr), HdR]
Henri Portine [University Bordeaux 3, Professor (Pr), HdR]

Technical Staff
Patrick Henry [CNRS, Research Engineer (IR)]

PhD Student
Pierre Bourreau [INRIA, CORDI/S, since Sep. 2007]
Bruno Mery [University Bordeaux 1, since Oct. 2006]
Benoît Razet [University Paris 7, located in Rocquencourt]
Émilie Voisin [University Bordeaux 3]

Post-Doctoral Fellow
Daniele Porello [INRIA]

Visiting Scientist
Denis Delfitto [University of Verona, INRIA Invited Professor (Pr), 15 March – 15 May 2008, HdR]

Administrative Assistant
Sylvie Embolla [INRIA]

2. Overall Objectives

2.1. Overall Objectives

The Signes project-team addresses several related domains of computational linguistics: :

- flexional and derivational morphology,
- syntax,
- logical semantics,
- lexical semantics,
- discourse representation.

The interfaces are of special interest to our group: between morphology and syntax, between syntax and semantics, between morphology and semantics, between semantics and discourse structure,...

There are various means to tackle these questions. In Signes, the methological accent is on the formal, symbolic methods issued from logic, which are also studied for themselves, like:

- formal language theory,
- categorial grammars,
- resource logics,
- lambda calculus,
- higher order logic.

We also develop the corresponding computational linguistics tools, which result in natural language processing (NLP) pieces of software for analysis, generation and acquisition devices. Some specific languages deserve particular attention, like Sanskrit, French Sign Language, Dutch and French:

- natural language tools for Sanskrit,
- modelling of French Sign Language grammar,
- large scale grammar for the NWO Dutch Spoken Corpus,
- lexicon and grammar for the analysis of written French.

3. Scientific Foundations

3.1. Word structure and automata for computational morphology

Keywords: *finite state automata*, *morphology*, *transducers*.

Participants: Gérard Huet [correspondent], Kim Gerdes.

Computational models for phonology and morphology are a traditional application of finite state technology [62], [63], [64], [44]. These models often combine symbolic or logical systems, like rewriting systems, and statistical methods like probabilistic automata which can be learnt from corpus by Hidden Markov Models [70].

Morphology is described by means of regular transducers and regular relations, and lexical data bases, as well as tables of phonological and morphological rules are compiled or interpreted by algebraic operations on automata.

The existing techniques for compiling such machinery are rather confidential, while any naive approach leads to a combinatorial explosion. When transformation rules are local, it is possible to compile them into an invertible transducer directly obtained from the tree which encodes the lexicon.

A generic notion of sharing allows to have compact representation of such automata. Gérard Huet has implemented a toolkit based on this technique, which allows a very efficient automatical segmentation of a continuous phonologic text.

This study of the linear structure of language and of word structures is by itself sufficient for applications like orthographic correctors and text mining. Furthermore, this preprocessing is required for the analysis of other layers of natural language like syntax, semantics, pragmatics, etc.

3.2. Sentence structure and formal grammars: syntax

Keywords: *categorial grammar*, *descriptive linguistics*, *formal grammar*, *formal languages*, *lambda calculus*, *metagrammar*, *monadic second order logic*, *natural language processing*, *parsing*, *resource logic*, *transducers*.

Participants: Lionel Clément, Kim Gerdes, Anna Kupść, Alain Lecomte, Renaud Marlet, Richard Moot [correspondent], Christian Retoré, Sylvain Salvati.

Sentence (or phrasal) structure is usually modelled via a tree structure. Different families of syntactic models are studied in Signes: rewriting systems of the Chomsky hierarchy, including tree grammars, deductive systems, i.e. categorial grammars, and constraint-based approaches.

3.2.1. *Formal Grammars*

Rewriting systems have excellent computational properties and a quite good descriptive adequacy. Relevant classes of grammars for natural language syntax, the so-called mildly context sensitive languages, are just a bit beyond context-free languages, and they are parsable in polynomial time as well well [58]. Among these classes of grammars let us mention Tree Adjoining Grammars [56], [57] and Minimalist Grammars [81], [82], [72]. Dependency Grammars and Lexical Functional Grammars share some properties with them but the general paradigm is quite different [71], [49].

Edward Stabler in [81] introduced Minimalist Grammars (MGs) as a formalization of the most recent model of the Chomskian or generative tradition and they are quite appealing to us. They offer a uniform model for the syntax of all human languages.

- There are two universal, language independent, rules, called merge and move: they respectively manage combination of phrases and movement of phrases (or of smaller units, like heads).
- Next, a language is defined by a (language dependent) lexicon which provides words with features describing their syntactic behavior: some features trigger merge and some others move. Indeed, features have positive and negative variants which must cancel each other during the derivation (this is rather close to resource logics and categorial grammars).

Consequently, MGs are able to describe numerous syntactic constructs, providing the analyzed sentences with a fine grained and complete syntactic structure. The richer the syntactic structure is, the easier it is to compute a semantic representation of the sentence.

MGs also cover phenomena which go beyond syntax, namely morphology via flexional categories, and they also incorporate some semantic phenomena like relations between pronouns and their possible antecedents, quantifiers, etc.

A drawback of rewrite systems, including MGs, is that they do not allow for learning algorithms that can automatically construct or enlarge grammars from structured corpora. But their main drawback comes from the absence of structure on terminals, which gives no hint about the predicative structure of the sentence.

Indeed, a strong reason for Signes using categorial grammars and their extensions [73]. Indeed, despite the inefficiency and the restricted linguistic coverage initial categorial grammars (BA, Lambek) provide a correspondence between syntactic analyses and semantic representations, which we are trying to extend to richer formalisms. This will be explained in the next section on the syntax/semantics interface.

In order to improve the computational properties of categorial grammars, and to extend their scope, we have been working on connecting them to more efficient and wider formalisms, like MGs [67], [66], [80].

A relatively new approach to syntax is known as model-theoretic syntax. Its advantages have been underlined by Geoffrey Pullum in [78]. Instead of viewing the trees or strings as a closure of some base set of expressions, they are viewed as trees or sets satisfying a set of formulae. This approach may be considered as another way of describing generative grammars. The advantages of such a description are not in the parsing algorithms (MSO or Constraint Satisfaction are usually of high complexity) but rather in characterising the language class and possibly describing it in a linguistically natural way (as opposed to lexical items of lexicalized grammars). This connection to logic is related to constraint-logic programming or to monadic second order logic.

In the MSO style, the pioneering work of James Rogers on Government and Binding and Tree Adjoining Grammars must be mentionned in [78]. Uwe Mönnich, Jens Michaelis and Frank Morawietz have obtained a two step description of minimalist grammars that we are studying [75], [74].

In the constraint style issued from the Prolog-Definite Clause Grammars, Head Phrase Structure Grammar, Construction Grammars and Property Grammars are defined as sets of constraints. The later ones introduced by Philippe Blache offer a rather natural way to describe grammar rules and have been studied by Marie-Laure Guénot in our group [50], [45].

3.2.2. *High-Level Syntactic Formalisms*

Lionel Clément worked on a formal representation of grammatical generalisations implemented for several linguistic formalisms.

This work deals with the problem of same linguistic phenomena expressed in several formalisms, alternative realisations and linguistic generalisations. The project aims at finding a common representation platform for all considered formalisms and factoring out elements shared in different linguistic constructions (i.e. different realizations of a nominal subject). The alternatives describe sets of related grammatical constructions (i.e. diathesis alternations). Finally, the shared part of these descriptions is expressed in a high-level linguistic formalism closely related to metagrammar representations [46], [47], [48]. For instance, diathesis alternations can be considered an intersection of syntactic realizations of passive, active or causative sentences.

As exposed on the ARC Mosaïque web site http://mosaique.labri.fr/, the new idea introduced in the metagrammar paradigm is the fact that metagrammars handle two kinds of factorized informations: structural (and formalism dependent: tree structures, graphs, dependancies), and linguistic. The latter presupposes introducing of a way to represent non generative data and linguistic knowledge, without redundancy.

3.2.3. *Linguistic Formalisations*

In addition to studying formal properties of the models mentioned above, Signes use them to describe linguistic phenomena in various languages. Dependency Grammars have been applied to a detailed analysis of word order in German, whereas various French phenomena have been formalised and implemented as computational grammars adopting Property Grammar frameworks [50], [45]. Lexical Functional Grammar: the XLFG parser implemented by Lionel Clément. Finally, a morphosyntacitc analysis of Polish past tense and conditional verb forms has been modelled in HPSG. This formalism has also been used by members of the group to account for French inflection.

3.3. Sentence structure and logic: the interaction between syntax and semantics

Keywords: *Montague semantics*, *categorial grammars*, *compositional semantics*.

Participants: Renaud Marlet, Richard Moot, Christian Retoré, Sylvain Salvati [correspondent].

The principle of compositionnality, enounciated by Frege and formalized by Montague, makes syntax be the main window through which the meaning of natural language can be studied. This is the reason why mathematical linguistics is not only interested in the kinds of languages formalisms may describe, but also to the structures those formalisms assign to valid sentences. In his work Montague uses the lambda calculus so as to build the semantics of fragments of English, and the connection with the lambda calculus that the Curry-Howard correspondence naturally establishes with many kinds of intuitionnistic logics has made categorial formalisms quite popular for developing Montague semantics. But Montague's work may also be adapted to other syntactic formalisms and it gives some rather elegant formalization of semantics [77].

In Montague's approach the syntactic structure of a sentence is also the structure of its meaning and thus a from a grammar *à la* Montague, one can extract a grammar of a language and a grammar of meaning. This leads to the study of languages of structures denoting meanings. In particular, as in Montague tradition, when meanings are represented as logical formulae, themselves represented as lambda terms, this leads to a study of languages of lambda terms.

Montague's approach is only a first step toward modeling semantics and it has several shortcomings. For example, the particular kind of transformation it uses to map syntactic structures to meanings sometimes require to model syntax in a way that is counter-intuitive. The truth-conditional and intentional models of meaning, even though they can explain some phenomena of entailment, are not relevant from a cognitive point of view [55]. It is also the case that the strict compositional principle does not hold in general, as the famous Geach examples shows.

The interests of the project at the interface between syntax and semantics are oriented in three directions. First, it seems that the minimalist approach of syntax advocated by Chomsky provides a description of language that is very close to it semantics. Thus we try to provide techniques for giving semantics to minimalist grammars. Second the various representations of discourse semantics are a way to extend Montague semantics and to make it more sensitive to the context. Finally, we try to understand the expressive power of Montague semantics so as to assess its possibilities.

3.4. Lexical semantics

Keywords: *computational semantics*, *lexical semantics*.

Participants: Christian Bassac, Mauro Gaio, Patrick Henry, Renaud Marlet, Bruno Mery, Christian Retoré [correspondent].

One of the most exciting challenges in computational linguistics is the question of lexical semantics, that is a proper treatment of word meanings and the way they relate one to another and finally how to handle the minimal interaction with knowledge representation. This part of semantics is relevant, not to say mandatory, for computing the semantic counterpart of composition be it lexical or syntactic.

The Generative Lexicon [79] is one of the most common frameworks for representing the internal structure of the meaning of words and morphemes, an alternative being the lexical functions of Mel'c̆uk and Polguère. The former is better suited for the logical apparatus developed by Signes, since Pustejovsky set up can be viewed as an extension of Montague semantics, with which it shares the compositionality and the type theoretical formulation.

The information which depicts the sense of a word or morpheme is organised in three layers: the argument structure (related to logical semantics and syntax), the event structure, and the qualia structure. The argument structure provides types (in the type-theoretical sense) to the arguments encoded in the qualia structure regardless of whether they are syntactically mandatory or optional. The event structure follows [60]. It unfolds an event into several ordered sub-events with a mark on the most salient sub-event. Events are typed according to the typology of Vendler: state, process and transition, this latter type including achievement and accomplishment. The qualia structure relates the argument structure and the event structure in roles: formal, constitutive, telic, agentive.

This information and its organization into the generative lexicons allows an explanation of, for instance, polysemy and of compositionality (in particular in compound words or in simple phrase structure). This kind of model relates knowledge representation to linguistic organization and thus is especially useful for word sense disambiguation during (automated) syntactic and for computing the semantics of a compound, a phrase, a sentences and a discourse.

Signes is for instance interested in the so-called logical polysemy, that is how some occurrences may refer to one or another aspect (corresponding to a semantic type) of a given word. In order to get a better interface with syntax, our research rather try to extend logical and compositional sentence-semantics like Montague semantics and lambda-DRT, than to encode the structure that one finds in dictionaries and lexical studies.

3.5. Discourse and dialogue structure: computational semantics and pragmatics

Keywords: *DRT*, *Montague semantics*, *SDRT*, *computational semantics*.

Participants: Mauro Gaio, Alain Lecomte, Renaud Marlet [correspondent], Christian Retoré.

Montague semantics have some limits. Two of them which, technically speaking, concern the context, can be overcome by using DRT, that is Discourse Representation Theory and its variants. [60], [84] Firstly, if one wants to construct the semantics of a piece of text, one has to take into account sequences of sentences, either discourse or dialogue, and to handle the context which is incrementally defined by the text. Secondly, some constructs do not obey the strict compositionality of Montague semantics, since pronouns can refer to bound variables. For instance a pronoun of the main clause can be bound in a conditional sub-clause.

For these reasons, Discourse Representation Theory was introduced. This model defines an incremental view of the construction of discourse semantics. As opposed to Montague semantics, this construction is top-down, and proceeds more like state change than like functional application — although lambda-DRT present DRT in a Montague style, see e.g. [84].

These approaches may be used for constructing semantic representations of fragments of natural language. Such representations are relevant for applications like information extraction and retrieval, question answering system, and human-computer interaction, among others.

4. Application Domains

4.1. Sanskrit philology

Keywords: *Indian studies*, *Internet*, *Sanskrit*, *natural language processing*.

Participant: Gérard Huet [correspondent].

Sanskrit literature is extremely rich, and is part of the world cultural patrimony. Nowadays, Internet can provide to both specialists and inquiring minds an access to it.

This kind of resource already exists for ancient Greek and Latin literature. For instance, Perseus (http://www.perseus.tufts.edu) provides an online access to texts. A simple click on each word analyses it, and brings back the lexical item of the dictionary, possible meanings, statistics on its use, etc.

The work described in the following sections enables such computational tools for Sanskrit, some of which are already developed and made available on a web site (http://sanskrit.inria.fr). These tools efficiently and accurately assist the annotation of Sanskrit texts. Besides, a tree bank of Sanskrit examples also is under construction. Such a corpus annotation tool is a prerequisite to the implementation of a Perseus-like facility for Sanskrit.

4.2. Towards French Sign Language (LSF) modelling and processing

Keywords: *deaf community*, *disabled*, *multimedia communication*, *sign language*.

Participants: Lionel Clément, Patrick Henry, Renaud Marlet, Henri Portine, Christian Retoré, Émilie Voisin [correspondent].

After a global prohibition of Sign Languages decided in 1880 (and which lasted until the sixties in the USA and until the eighties in France), deaf people can use sign language and rather recently these languages are the object of new studies and development. A first aspect is the social acknowledgment of sign language and of the deaf community, a second aspect is the linguistic study of this language with a different modality (visual and gestural as opposed to auditive and phonemic), and the third and most recent aspect, which relies on the second, is the need for sign language processing. A first goal is computer aided learning of Sign Language for hearing people and even deaf people without access to sign language. A more challenging objective would be computer aided translation from or to sign language, or direct communication in sign language.

Given the rarity of linguistic study on the syntax and semantics of sign languages — some exceptions concerning American Sign Language are [76], [68], [69] — before being able to apply our methodology, our first task is to determine the structure of the sentence, using our personal competence as well as our relationship with the deaf community.

We intend to define methods and tools for generation of sign language sentences. It should be noted that there is a continuum of different representations of a sentence in Sign Language, from a grammatical description with agreement features and word/sign order that we are familiar with, to a notation system like Signwriting [83] or to a language for the synthesis of 3D images and movies. Our competences on the interface between syntax and semantics are well designed for work in generation of the grammatical representations.

A first application would be a software for teaching Sign Language, like the CD ROM *Les Signes de Mano* by IBM and IVT. Indeed, presently, only dictionaries are available on computers, or examples of sign language videos, but no interactive software. Our generation tools, once developed, could be useful to educative purposes.

4.3. Parsers

Participants: Lionel Clément [correspondent], Renaud Marlet, Richard Moot, Sylvain Salvati.

In the implementation of a robust parser, one of the major issue arises from homonymous words and phrases. Natural language is highly ambiguous and each sentence, taken without any pragmatic or semantic context, has a huge number of possible meanings. In written languages this combinatorial problem necessitates the use of subtle techniques; but in spoken languages, where normative rules have less influence, those techniques do not seem to be able to cope with ambiguity. The recent developments of natural language processing concerning the problem of ambiguity is based on stochastic and low-level methods. Those techniques try only to represent surface dependencies and forget about the various structures of phrases and about their meanings. They are quite efficient for applications such as information retrieval and lack accuracy in others like automatic translation.

We would like to develop new techniques so as to allow robust parsing of spoken language, but also so as to deal with the computation of meaning regardless the ambiguity of sentences. Usually the various possible analyses of a sentence are represented in a structure called "shared forest". Such a structure can be seen as a tree automaton. This remark gives us several directions of research. A first one would be to adapt various techniques coming from automata theory especially concerning automaton transformations and transductions. A second one consists in using the connection between tree automaton theory and the weak MSO theory of trees so as to perform selections of certain sets of analyses.

4.4. Syntactic and semantic modeling

Participants: Lionel Clément, Renaud Marlet, Richard Moot [correspondent], Sylvain Salvati.

The SIGNES team develops Grail, for multimodal categorial grammars, and XLFG, for Lexical-Functional Grammars as well as several different grammars with different levels of grammatical coverage for these formalisms, ranging from the specific - giving descriptions of linguistic phenomena such as French clitics and extraposition - to wide-coverage: a wide-coverage categorial grammar of Dutch has been developed while we intend to develop a similar wide-coverage grammar for French.

Categorial grammars have a transparent syntax-semantics interface by means of the Curry-Howard homomorphism, where a proof of the grammaticality of a sequence of typed words immediately gives us the way of combining the corresponding semantic expressions. It is our goal to develop this correspondence to allow for wide-coverage semantic analysis, using Pustejovsky's generative lexicon to help with semantic disambiguation.

A specific application we envisage, collaborating with researchers from Pau and Toulouse, is to perform syntactic, semantic and discourse analysis of a corpus reciting voyages through the region of the Pyrnes. Naturally, this chain of analysis will be specialized towards the application domain, benefiting from specific knowledge concerning the region, means of transport and the conventions of this style of literature.

5. Software

5.1. The Zen Toolkit

Keywords: *computational morphology*, *finite state technology*, *functional programming*, *natural language processing*, *segmentation*.

Participant: Gérard Huet [correspondent].

The Zen Toolkit is a library of finite state automata and transducers, called Zen for its simplicity. The algorithmic principles of the Zen library are based on the linear contexts data structure ("zippers") and on the sharing functor (associative memory server) [51]. It has been developed by Gérard Huet and is being used in his Sanskrit modelling platform (see section 5.2). It allows the construction of lexicons, the computation of morphological derivatives and inflected forms, and the segmentation analysis of phonetic streams modulo euphony [51].

The Zen Toolkit is implemented in an applicative kernel of Objective Caml, called Pidgin ML. It follows a *literate programming* style of documentation, using the program annotation tool Ocamlweb of Jean-Christophe Filliâtre, available for Ocaml. The Zen toolkit is distributed as free software (under the LGPL licence) in the Objective Caml Hump site as well as at URL http://sanskrit.inria.fr/ZEN/. This development forms a significant symbolic manipulation software package within pure functional programming, which shows the feasibility of developing in the Ocaml system symbolic applications having good time and space performance, within a purely applicative methodology.

The Zen Toolkit has been used, e.g., to implement a lexicon of french flexed forms (Nicolas Barth and Sylvain Pogodalla, Calligramme project-team at Loria). It is also used by Aarne Ranta (Chalmers University) as a morphological engine of the Grammatical Frameworks software.

5.2. Sanskrit Site

Keywords: *Sanskrit*, *electronic dictionary*, *parsing*, *segmentation*, *tagging*.

Participant: Gérard Huet [correspondent].

Gérard Huet's Sanskrit Site (http://sanskrit.inria.fr) provides a unique range of interactive resources concerning Sanskrit philology [53], [52]. These resources are built upon, among other ingredients, the Zen Toolkit (see section 5.1). The site registers thousands of visitors daily.

- The *declension engine* gives the declension tables for Sanskrit substantives.

- The *conjugation engine* conjugates verbs for the various tenses and modes.

- The *lemmatizer* tags inflected words.

- A *dictionary* lists inflected forms of Sanskrit words. Full lists of inflected forms, in XML format (given with a specific DTD), are released as free linguistic resources available for research purposes. This database, developed in collaboration with Pr. Peter Scharf, from the Classics Department at Brown University, has been used for research experiments by the team of Pr. Stuart Shieber, at Harvard University.

- The *Sanskrit Reader* segments simple sentences, where the (optional) finite verb form occurs in final position. This reader enhances the hand-tagged Sanskrit reader developed by Peter Scharf, that allows students to read simple texts differently: firstly in davanagari writing, then word-to-word, then in a word-to-word translation, then in a sentence-to-sentence translation.

- The *Sanskrit Parser* eliminates many irrelevant pseudo-solutions (segmentations) listed by the Sanskrit reader.

- The *Sanskrit Semantic Analyzer*, based on the notion of *kāraka* of Pāṇini, controls overgeneration using a pertinence principle [54].

- The *Sanskrit Tagger* is an assistant for the tagging of a Sanskrit corpus. Given a sentence, the user chooses among different possible interpretations listed by the morpho-syntactic tools and may save the corresponding unambiguously tagged sentence on disk as an hypertext document indexing in the Sanskrit Heritage Dictionary (our structured lexical database). This service has no equivalent worldwide.

- The *morphological data* for Sanskrit have been released by Gérard Huet under LGPLLR (http://sanskrit.inria.fr/DATA/XML/). The precise lexer used by the shallow parser is specified as a *modular transducer* whose top-level states are the lexical categories corresponding to the flexed forms banks, and whose arcs correspond to (the inversion of) euphony (*sandhi*) rules.

An on-going project is the construction of a tree bank of Sanskrit examples, in collaboration with Pr. Brendan Gillon, from McGill University in Montreal.

5.3. Grail: Natural Language Analysis with Multimodal Categorial Grammar

Keywords: *logic programming*, *parsing*, *semantic analysis*, *syntactic analysis*.

Participant: Richard Moot [correspondent].

Grail is a modern, flexible and robust parser/automated theorem prover for multimodal categorial grammars (MMCG [73]) developed by Richard Moot. It is designed to allow students and researchers to design and experiment with their grammars while at the same time offer the advanced users many optional optimisation strategies.

Grail can be run either as a command line script or as an application with a graphical interface, with the possibility to follow Grail's partial parses/proof attemps interactively. It is feely available from http://www.labri.fr/perso/moot/grail3.html.

5.4. Suite of Corpus Tools for Type-Logical Grammars

Keywords: *corpus*, *type-logical grammar*.

Participant: Richard Moot [correspondent].

A suite of corpus tools has been developed by Richard Moot. It contains tools for the display, search, transformation and extraction of grammars on the base of an annotated corpus. In addition, there is are tools for the generation of training and test data for maximum entropy models, a supertagger and scripts for error analysis included in the tools. Grail and the supertagger are designed to work in tight integration. This suite of corpus tools is available from http://www.labri.fr/~moot/Corpus/.

5.5. XLFG5: Experimental LFG Parsing

Keywords: *Lexical Functional Grammar*, *parser*.

Participant: Lionel Clément [correspondent].

XLFG5 is a parser prototype for research. It implements a variant of the Lexical Functional Grammar (LFG) formalism. The parsing produces a shared forest (of c-structures) in order to speed-up ambiguous sentences analyzes. Sharing functional dependency structures (f-structures) is under development. XLFG5 has been developed by Lionel Clément and is available as a parsing server (http://www.xlfg.org/).

XLFG5 is used for teaching in various universities, amongst which Université Bordeaux 3, Université Paris IV, and others in Spain or Algeria. Languages known to be used with XLFG5 are french, arabic, mandarin, spanish, english, german and thai.

5.6. Yab: a GLR Parser Generator for S-Attributed Grammars

Keywords: *compiler compiler*, *parsing ambiguities*, *parsing sharing*.

Participant: Lionel Clément [correspondent].

YAB is a GLR parser generator for S-Attributed grammars. This compiler has been used to develop a syntactic parser dealing with homonymies in LFG. It relies on a specific restriction of the LFG formalism to build a polynomial-time syntactic parser. This software has been developed by Lionel Clément (before he joined the Signes project-team). It is publicly available (http://www.labri.fr/perso/clement/yab/).

5.7. Lexed: a Dictionary Lexicalizer

Keywords: *dictionary search*, *lexicalizer*.

Participant: Lionel Clément [correspondent].

Lexed is a lexicalizer. It allows one to search a dictionary entry for a string. The finite automata-based algorithm is particularly fast, and offers a good alternative to hashes for large dictionnaries. Lexed is a C++ library distributed with a GPL Licence (http://www.labri.fr/perso/clement/lexed/). This software has been developed by Lionel Clément (before he joined the Signes project-team).

5.8. TreeLex: a Subcategorisation Lexicon

Keywords: *adjective*, *lexicon*, *treebank*, *valence*, *verb*.

Participant: Anna Kupść [correspondent].

TreeLex is a subcategorisation lexicon automatically extracted from a syntactically annotated corpus (tree-bank). It contains about 2000 contemporary French verbs (types) with their valency frames. We have identified 180 different verb frames, with 2.09 frames per verb on average. We used a mixed representation, i.e., both the function and a syntacitic category of an argument are specified. Recently, the work has been extended to the extraction of adjective frames. TreeLex contains 2153 qualitative adjectives (types) with 41 different frames discovered in the corpus. The vast majority of adjectives appear solely with a nominal subject and only 304 adjectives were found with a different or an additional frame.

The lexicon has been created and is maintained by Anna Kupść (the work started before she joined Signes). It is freely available at http://erssab.u-bordeaux3.fr/article.php3?id_article=150.

5.9. DepLin: Text Generation from Syntactic Dependencies

Keywords: *natural language generation*, *syntactic analysis*, *syntactic dependency*, *topological grammar*.

Participant: Kim Gerdes [correspondent].

DepLin takes a syntactic dependency tree as the input. The topological grammar translates such an (unordered) tree to an ordered constituent tree, called a topological tree. In the following step, this tree is simplified to a three level prosodic constituent tree (prosodic words, prosodic phrases, prosodic sentences). From this tree, a very simple sound output device can concatenate prerecorded sound files corresponding to the different prosodic words (with their prosodic markup). This allows for auditory tests of the resulting sentences in constructed communicative contexts (question-answer sets). The construction of the prerecorded files is quite time consuming; it has been tested on small vocabulary of Modern Greek.

DepLin was developed by Kim Gerdes. It is distributed as free software under GPL (http://gerdes.fr/soft/deplin/) and, apart from its use in Signes (in particular for German and Greek), it is mainly used at the University of Paris 7 for the development of different grammars (in particular Arabic and French).

5.10. Alignator: Bilingual Text Alignment

Keywords: *bilingual text alignment*, *textometric analysis*.

Participant: Kim Gerdes [correspondent].

A preliminary version of a bilingual text alignment tool, called Alignator, has been developed made available (http://elizia.net/alignator). It is written in Python, C, and Javascript, It runs on a web server for high accessibility. It aims at being a useful tool for researchers working on bilingual text who need fast paragraph alignment for textometric analyses (cf. section 6.4.1).

5.11. GrosMoteur: Acquisition of Linguistic Corpora on the Web

Keywords: *corpora*, *web-robot*.

Participant: Kim Gerdes [correspondent].

Kim Gerdes is developing a web-robot specialized in linguistic corpora acquisition with automatic domain and language recognition and recognition of morphological schemas for inflected languages. (See also section 6.4.2.) It is available from http://grosmoteur.elizia.net/.

5.12. French Sign Language HD Corpus

Keywords: *French Sign Language*, *LSF*, *coprus*, *high definition*.

Participants: Patrick Henry [correspondent], Émilie Voisin.

Thanks the two HD-video cameras provided by a former regional grant, Émilie Voisin and Patrick Henry recorded two hours of sign language utterances from native deaf speakers. This corpus is made of translations of French texts and short comics strips (with no textual annotation) turned into sign language expressions; it also contains sign language productions by non-native speakers as well as the reactions and comments of native sign language speaker confronted to these productions. Annotation and extraction of sequences in the corpus can be performed thanks to a specific interface.

5.13. Other On-going Software Developments

In the following, we list various on-going software developments. Most of these software and resources are in an early stage, and not yet packaged or available. Please get in touch with their correspondent to know the current status.

5.13.1. Hyperion: a parser for HR grammars

Keywords: *hyperedge replacement grammar*.

Participant: Richard Moot [correspondent].

Hyperion is a hypergraph parser developed by Richard Moot. It analyses graphs using hyperedge replacement grammars. The core parser is a very small and very general implementation of Clemens Lautemann's dynamic programming algorithm for parsing graphs using hyperedge replacement grammars. A rudimentary Dot/Graphviz interface for portraying the hypergraphs and grammars is included. Hyperion is written in XSB Prolog.

Hyperion is a first implementation of many of the author's ideas which were presented at TAG+9 [32] as well as in talks in Bordeaux and Chieti (cf. §8.6.2). An early version of the system can be dowloaded from http://www.labri.fr/perso/moot/hyperion/.

5.13.2. Datalog parser for PMCFG

Keywords: *parallel multiple context-free grammar*.

Participant: Sylvain Salvati [correspondent].

Sylvain Salvati has developed a recognizer for PMCFG (Parallel Multiple Context Free Grammars) which has the correct prefix property. This recognizer is based on a compiler that transforms a grammar (a PMCFG) into a datalog program which in turn is optimized in several steps. Recognition is reduced to solving a query with this datalog program, the analyzed sentence being encoded as an extensional database. A datalog query solver is also provided with this prototype. The optimization of the datalog program is based on a modular approach which composes several atomic transformations. The composition of these transformations is enforcing properties of the recognition algorithm. Therefore composing different transformations yields to different recognition algorithms. This program is based on an extension of Kanazawa [61] prefix correct algorithm for MCFG (Multiple Context Free Grammars) proposed by Sylvain Salvati. This program is already used by Kanazawa and by the Calligramme INRIA team-project in Nancy. In the future it should be improved in several ways; first it should become a parser; second it should be interfaced with certain classes of Abstract Categorial Grammars, so as to perform text generation; finally, so as to face the scaling problem of handling big grammars, compositions of transformations should be compiled into one efficient transformation.

5.13.3. *Farsi Corpus*

Keywords: *Farsi*, *Persian*, *annotation*, *corpus*.

Participant: Kim Gerdes [correspondent].

In collaboration with Pollet Samvellian (University Paris 3), Kim Gerdes developed an automatically annotated large corpus of Farsi (see section 6.4.2). This software will be distributed as a part of an associated project with the University of Berlin.

5.13.4. *Generative Lexicon Toolkit*

Keywords: *Generative Lexicon*, *anaphoric reference*, *network*, *nominal compound*.

Participants: Christian Bassac, Patrick Henry [correspondent].

Patrick Henry and Christian Bassac designed a toolkit for the implementation of a Generative Lexicon which can be shared and used on a network. This tool is in particular designed to filter anaphoric reference in nominal compounds. The implementation of other features of the Generative Lexicon is planned.

5.13.5. *Turkish Conjugation*

Keywords: *Turkish conjugation*.

Participants: Christian Bassac, Patrick Henry [correspondent].

Christian Bassac and Patrick Henry started working on the development of a platform designed to implement the generation of verb forms of Turkish conjugation paradigms. Some work remains to get done mainly regarding the details of the features of number and person, and on the labelling of forms. http://signes.bordeaux.inria.fr/turc/turc.html

6. New Results

6.1. Properties of the Formalisms

6.1.1. *Survey on the Mathematics of Computational Linguistics*
Participant: Christian Retoré [correspondent].

Christian Retoré wrote a sixty pages survey for mathematicians on the mathematics for computational linguistics. This survey is divided into two parts: the first one is devoted to formal language theory [23] and the second to logic [22]. They both mainly focus on syntax, because the mathematics of semantics are up to now not so mature. It includes some history of the models and a personal view on their evolution, as well as the tense relation bewteen the linguistic view and the applicative one. The first part presents string and tree languages and their linguistics use, and what is known about the languages class needed for language description, including a few words on the Mönnich, Michaelis, Morawietz two step approach — see also the book review he wrote in TAL on Morawietz's book [18]. The second part focuses on type logical grammars, with some mathematical results on Lambek calculus including a yet unpublished result with Pogodalla on proof nets, an up to date account of learning, and Montague semantics for such formalisms. It ends with the accurate question of linking traditional formal language theory and categorial syntax as explored by Kanazawa, Salvati, Retoré, Stabler and Kobele, and a discussion of the possible research directions in linguistically motivated mathematics.

6.1.2. Relational Machines

Participants: Benoît Razet [correspondent], Gérard Huet.

Benoît Razet continued his doctoral research on Eilenberg machines in the framework of functional programming. Such machines provide a general paradigm for non-deterministic programming, encompassing standard automata theory notions. They are appropriate to the modeling of natural language processing, specially in its finite-state aspects. He defined a restriction of these machines to a notion of "finite machine" which limits the search space to a finite exploration. For such machines, a complete simulation may be obtained by depth-first search, in line with the reactive engine issued from the Zen toolkit. Furthermore, this simulation is efficient, since it may be implemented by a flowchart (i.e. recursive calls are terminal). The framework of finite Eilenberg machines was presented by Benoît Razet at the CIAA conference in San Francisco in July 2008 [33].

One of the benefits of using functional programming for the specification of such computing processes is to replace the usual paraphernalia of automata (tapes, reading heads, stacks, etc) by clear mathematical concepts. Thus finite Eilenberg machines may be axiomatized in a formal proof assistant such as Coq, the proof of soundness and completeness of their simulation by the reactive engine may be expressed in such a proof system, and the reactive engine program may be mechanically extracted from the formal development, leading to a formally certified computation paradigm. This was presented by Benoît Razet at the MSFP workshop, within the ICALP conference in Reykjavik in July 2008 [34].

More general strategies may be used as parameters to a functorial description of such relational machines. The control component of the machines may be compiled from regular expressions. Benoît Razet studied several such algorithms, and succeeded in their applicative implementation. A tutorial on this methodology of computing with relational machines was presented by Gérard Huet and Benoît Razet at ICON'2008 [40].

6.1.3. Type-Theoretical grammars and proof theory

Keywords: *Abstract Categorial Grammar*, *Lambek calculi*, *Linear Logic*, *Minimalist Grammars*, *membership problem*, *mildly context sensitive formalisms*, *parsing*, *phenogrammar*, *tectogrammar*.

Participants: Pierre Bourreau, Richard Moot, Christian Retoré, Sylvain Salvati [correspondent].

Richard Moot shows that two recent extension of the non-associative Lambek calculus, the Lambek- Grishin calculus and the multimodal Lambek calculus generate class of languages as tree adjoining grammars. To obtain this result he uses (tree generating) hyperedge replacement grammars as an intermediate step. As a consequence both extensions are mildly context-sensitive formalisms and benefit from polynomial parsing algorithms.

Daniele Porello and Christian Retoré studied the logic PCL (partially commutative logic) developed by de Groote (1996) and the later one in order to provide proof nets for PCL extending the the work by Christian Retoré on proof nets without links for commutative multiplicative linear logic (2003) and his extension to cyclic multiplicative linear logic obtained with S. Pogodalla (2005).

Sylvain Salvati [20] has showed that the membership problem for second order non-linear abstract categorial grammars was decidable. An interesting side-effect of this result is the generation problem is decidable in systems using derivation trees and Montague semantics.

Christian Retore and Sylvain Salvati [19] propose an encoding of the non-associative Lambek grammars within the ACG framework.

Pierre Bourreau and Sylvain Salvati investigated extensions of λ-grammars which can be parsed and generated with Datalog programs.

Arhtur Ball and Sylvain Salvati have proposed studied an extension of Kanazawa's prefix correct algorithm for parsing MCFG. This extension deals with PMCFGs. It is the subject of Arthur Ball's master report [39].

6.2. Semantics

6.2.1. *Semantics of Minimalist grammars*
Participants: Alain Lecomte [correspondent], Christian Retoré.

Alain Lecomte [29] has enriched the proposal of Amblard [41] by providing a detailed account of the semantics of VPs, showing that the main constructionnalist theses (like Kratzer [65]) and the Theta theory [43] may be incorporated into a Minimalist Grammar.

6.2.2. *Computational Semantics with Rich Lexical Semantics*
Participants: Christian Bassac, Bruno Mery, Christian Retoré [correspondent].

Christian Bassac, Bruno Mery and Christian Retore continued the formalization of lexical semantics in an integrated logic framework. In [16], they survey the many approaches undertaken for lexical disambiguation in context, and then detail their own model in second-order logic of some points of interest of the Generative Lexicon Theory. In particular, they tackle the problem of dot objects, using the multiplicative conjunction of linear logic as a generic type for dot-objects without systematic projections, yet with a compound type related to its components.

6.2.3. *Correspondance between semantic graphs and logic*
Keywords: *Meaning-Text Theory (MTT)*, *Minimal Recursion Semantics (MRS)*, *logic*, *quantification*, *semantic graph*.

Participant: Renaud Marlet [correspondent].

Analysis (from text to meaning) and generation (from meaning to text) often rely on two different kinds of meaning representations: logical formulas and semantic graphs. These two tasks and semantic formulations belong to different traditions in linguistics, logic and computer science, with little interconnections. Renaud Marlet has analysed the advantages and drawbacks of each kind of representation. Instead of opposing them, as is often the case, he proposed to articulate them, giving a logical meaning to semantic graphs. For this, he studied the lack of precision in existing semantic graphs and the different kinds of information that can be supplemented to single out a precise meaning. Given such an additional information (although possibly underspecified), Renaud Marlet also proposed a translation from a semantic graph to a logical formula of Minimal Recursion Semantics (MRS). This translation is very general and covers cases of multiple predications over several entities, higher order predication and modalities [31].

6.2.4. *Semantic Zeugma and Coordination*
Participant: Lionel Clément [correspondent].

Lionel Clément studied particular semantic zeugma that are pictured by the following examples:

1. Tout jeune Napoléon était très maigre et officier d'artillerie / plus tard il devint empereur / alors il prit du ventre et beaucoup de pays (Prévert)

2. Son corps nageait dans l'eau verte, et son esprit dans l'opulence (Troyat)

3. [...] en criant sur les toits et sur l'air des lampions (Brassens)

4. Cet homme marchait loin des sentiers obliques/Vêtu de probité candide et de lin blanc (Hugo)

5. enfermée dans sa chambre et dans sa surdité (Roger Martin du Gard)

6. J'ai traversé sur mes souliers ferrés le monde et la misère (F. Leclerc)

He formalized semantic zeugma in Lexical Functional Grammar (LFG) but did not treat gapping or ellipsis. This formalization relies on the *mapping* operation in LFG which consists in realizing the syntactic function of each argument of a semantic predicate. It is also based on a new operator that allows to give a new instance of the predicate, *i.e.* the predicate itself but also the relation that it has with each of its semantic arguments. This operation uses *mapping* in order to pass the correct semantic arguments to the predicates.

6.2.5. *Turkish Morphology and Semantics*

Participant: Christian Bassac [correspondent].

Christian Bassac contributes a paper entitled *Factorisation des affixes de pluriel et de possessif en turc* [15] with M. Ciçek from the university of Gaziantep, for the special issue of the *Revue de Sémantique et de Pragmatique* on *The semantic aspects of coordination*, in which he shows that the relevant factor to account for the suspension of plural and/or possessive suffixes in coordinated NPs constructions is the semantics of the conjunct NPs: for the suspension of the possessive suffix to be possible, both Ns must be individual level Ns, or relational Ns, as in both cases the identifying predicate is encoded in the Telic Role of the qualia structure.

6.2.6. *Pragmatics*

Keywords: *argumentation*, *fallacies*, *ludics*, *presupposition*.

Participant: Alain Lecomte [correspondent].

Alain Lecomte, with his Marseilles' colleagues Myriam Quatrini, Marie-Renée Fleury and Samuel Tronçon, showed how ludics is fruitful for expressing language games (in Wittgenstein's sense) because of the power it has to represent a statement in the same time as a "proof" and as a "strategy in a game". He also proposed a new conception of *meaning*, based on interaction according to which the meaning of an utterance is the set of all the sentences which react the same way with regards to other utterances.

6.2.7. *Spatio-temporal Semantic Analysis*

Keywords: *corpus studies*, *geographic information system*, *indexing*, *information retrieval*, *lexical semantics*.

Participant: Mauro Gaio [correspondant].

The research here reported aims at developing a toolset that end-users could use, first to retrieve travel stories from a large set of local cultural heritage corpora, and then to study the itineraries reported in these travel stories. To provide an adequate support to end-users, we propose two computational models from which we have built a Geographical Information Retrieval toolset in tune with travel stories characteristics.

In [30], Loustau Pierre, Thierry Nodenot and Mauro Gaio present some computational models and a toolset that were designed to address some needs of specific end-users like teachers trying to make use of particular localized documents called "travel stories". They focus on the toolset required during the preliminary design steps of a Computer-Aided Learning (CAL) application which aims at teaching about a high level geographical information (travel itineraries) embedded in a travel story / a set of travel stories. In [17] the authors present an already developed and tested prototype of such CAL applications. In case of a use requiring an unguided interaction (like in tourism), the GIR phase can only be based on the processing of an extended noun syntagm gravitating around a toponym; in a previous work the authors have called that a Spatial Feature (SF). On the other hand, if the expected use requires a more guided interaction (i.e. teaching activities), then a more elaborate interpretation and an extra effort towards formalization are necessary. Here, GIR processing must allow one to discover a pattern of a granularity that is greater than a syntagm. The extraction of these patterns raises a discourse problem. From the automatic processing standpoint, the aforementioned uses require a phase of Information Extraction (IE) and one of Information Retrieval (IR). In [21], because of the field of their application, the authors talk about Geographic IE (GIE) and Geographic IR (GIR). They have promoted a chains for an automatic semantic tagging of itinerary aspects in order to obtain more realistic geographical information within textual documents.

Some of the presented techniques derive from discipline of Geographic Information Systems (GIS). In [35], Christian Sallaberry, Mauro Gaio, Damien Palacio and Julien Lesbegueries, in order to improve the GIR processes, focus on a better computational interpretation (i.e. modeling and reasoning) of spatial relationships for a more accurate approximation of the corresponding footprints. First, the authors propose to better integrate the way that space and the things in it are schematized in natural language. Then, they propose better suited GIS functions to approximate such relationships in order to infer more convenient footprints.

Finally in [36] Nicholas Asher, Philippe Muller and Mauro Gaio argue that spatio-temporal primitives are crucial in giving a full view of the spatial and temporal structure of texts. They think that temporal and spatial structure are projections of a more complex and more complete spatio-temporal structure. They make their case based on an analysis of movement verbs, showing how they contribute in an important way to both temporal and spatial structure within discourse.

6.2.8. *Discourse and semantic graphs*

Keywords: *DRT*, *semantics*.

Participant: Maxime Amblard [correspondent].

In collaboration with Johannes Heinecke and Estelle Maillebuau, Maxime Amblard proposed an extension of the formal representations of semantics used in the natural language processing tool developed at Orange Labs [24]. They covered issues relating to the building of semantic representations, in a linguistic framework. To obtain more detailed representations of the argumental structure of statements, they included insights from the DRT into the representational system, which is based on semantic graphs, to give an account of scope.

6.3. Morphology and Syntax

6.3.1. *French Sign language*

Keywords: *French Sign Language (LSF)*, *Generative Grammar*, *language typology*, *morphology*, *syntax*.

Participants: Patrick Henry, Renaud Marlet, Henri Portine, Christian Retoré, Émilie Voisin [correspondent].

Émilie Voisin and Maxime Amblard presented a communication about a formalisation of the French Sign Language syntax at the TALS workshop in Avignon [42]. This paper is about verbal inflection and sign order in French Sign Language. They study the effect of verbal inflection on sign order (for example SOV or OSV). They analyse this inflection like noun incorporation. The phenomenon is modeled using Minimalist Grammars.

Émilie Voisin completed her PhD on verb in French Sign Language. She showed that word order in French Sign Language is constrained by the type of the verb. She proposed a typology to classify verbs in french sign language and gave a characterization of the type of inflection. There are two kinds of inflections: proform inflection and locus inflection. Émilie Voisin worked on proform inflection and tried to show that it can be compared with a phenomenon like noun incorporation.

6.3.2. Sanskrit Processing

Participant: Gérard Huet [correspondent].

Gérard Huet continued his work on the computational treatment of Sanskrit. He presented his work at the 2nd International Symposium on Sanskrit Computational Linguistics organized by Brown University in May 2008. The final version of the paper will appear in a special volume "Topics in Sanskrit Computational Linguistics" published by Springer-Verlag [28].

Gérard Huet continued his collaboration with Prof. Amba Kulkarni from Hyderabad University on Sanskrit Computational Linguistics, in the framework of their "associated team" *Sanskrit* (http://yquem.inria.fr/~huet/ EA/Sanskrit08.html). In particular, a study of consistency of morphological databases obtained by different methods is under way. The next happening in this collaboration is the organization of the Third International Symposium on Sanskrit Computational Linguistics in Hyderabad in January 2009.

6.4. Corpus

6.4.1. Bilingual Text Alignment

Keywords: *Chinese*, *French*, *bilingual text alignment*, *corpus*.

Participant: Kim Gerdes [correspondent].

Kim Gerdes proposes a work that picks up an old idea of using a dynamic time warping algorithm in the bitext alignment process if neither dictionaries nor cognates are available. He gives improvements on the underlying distance metrics which gives astonishingly good results for quite literal translations of long texts with some insertions like introductions and translator's notes. In [25] he tackles the statistical and computational problems and in [26] he takes a more linguistic perspective on the language internal bases and the usefulness of the approach.

In a collaboration with Miao Jun, Kim Gerdes describes the translation theoretical usage of an online tool they developed and that allows for a graphic representation of search results in aligned bitexts [59].

6.4.2. Ressource-Free Corpus Construction and Partial Annotation of Light Verb Constructions for Farsi

Keywords: *Farsi*, *Persian*, *annotation*, *corpus*, *learning algorithm*.

Participant: Kim Gerdes [correspondent].

In collaboration with Pollet Samvelian [27], Kim Gerdes has studied the linguistic bases of Persian light verb constructions. The have showed a corpus based construction of lists of collocates for some common Persian verbs. The methods of corpus construction that they propose are language independent and give good results on a relatively small corpus of 20 million words. This confirms the power of association measures based on the hypergeometric distribution. The resulting lists show a graduation of lexicalization and the semantic homogeneity of some light verb subcategorization schemes which could be the reason for their wide usage.

6.4.3. Adjectival Frame Extension to the TreeLex Corpus

Keywords: *TreeLex*, *adjectival frame*, *corpus*, *lexicon*.

Participant: Anna Kupść [correspondent].

Anna Kupść has extended TreeLex (cf. 5.8) with adjectival frames. TreeLex is a syntactic lexicon for French automatically extracted from a treebank. Due to properties of the corpus (relatively small size but rich syntactic annotations), the extraction method was guided by linguistic knowledge rather than statistical techniques. In particular, it has been sought to relate arguments of adjectives with regular and productive constructions.

7. Other Grants and Activities

7.1. National research programs

7.1.1. *Groupement de Recherche C.N.R.S. 2521 Sémantique et modélisation*

Participants: Maxime Amblard, Christian Bassac, Pierre Bourreau, Patrick Henry, Alain Lecomte, Renaud Marlet, Bruno Mery, Richard Moot, Henri Portine, Christian Retoré [correspondent], Sylvain Salvatti, Émilie Voisin.

Signes is one of the fifteen research teams involved in the Groupe de Recherches CNRS 2521, GDR *Sémantique et Modélisation*, directed by Francis Corblin (Université Paris IV) 2002-2005, 2005-2008. This research program is divided into five "operations": modèles et formats de représentation pour la sémantique; les modèles à l'épreuve des données; sémantique et corpus; Les interfaces de la sémantique linguistique; sémantique computationnelle. The Signes project-team is part of the latter two operations (interfaces of linguistic semantics and computational semantics). See http://semantique.free.fr/ for details.

7.1.2. *ANR GEONTO Programme Masse de Données et COnnaissances*

Participant: Mauro Gaio [correspondent].

The GEONTO project is a 3-year project (2008-2010) funded by the French National Research Agency (ANR) involving 4 partners: LRI, LIUPPA, COGIT, IRIT. It focuses on interoperability of diverse data related to geographic information. The first part of the project consists in building several geographic ontologies reflecting several (*different* might be better) points of view. In order to complete these objectives, various approaches relying on techniques taken from natural language processing will be used. The second part of the project will study the alignment of ontologies built in the previous part. See http://geonto.lri.fr/ for details.

7.1.3. *ANR blanche PRELUDE*

Participants: Maxime Amblard, Alain Lecomte, Bruno Mery, Richard Moot, Sylvain Salvati, Christian Retoré [correspondent].

Signes is part of the national research program PRELUDE *Towards a theoretical pragmatics based on ludics and continuations* is a 3-year project (2008-2011) funded by the French National Research Agency (ANR) (November 2006 – November 2009) directed by Alain Lecomte (team *Structures Formelles de la Langue*, Univ. Paris 8). Other partners are the LaBRI (most of the Signes members from the LaBRI), the INRIA project-team Calligramme (LORIA, Nancy) and the Institut Mathématique de Luminy (Marseille). See http://anr-prelude.fr/ for details.

7.1.4. *ANR project Rhapsodie*

Participants: Lionel Clément, Kim Gerdes, Renaud Marlet [correspondent].

Rhapsodie is a 4-year project (2008-2011) funded by the French National Research Agency (ANR). It aims at building an annotated corpus of spoken French that will be freely distributed within the research community. This corpus will contain several kinds of spoken discourses, and its annotations will focus both on prosody and syntax, including topology and syntactic dependency. It will be a precious resource to understand the status of prosody in spoken French and its relation with syntax and informational structure. It will also be a suitable resource to train probabilistic parsers targetted at spoken French. The project includes partipants from several research groups in France, as well as Belgium and Switzerland. The involvement of Signes members in Rhapsodie mainly concerns syntax. See http://rhapsodie.risc.cnrs.fr for details.

During the 1st year of the project, the Signes participants have analysed existing data and contributed to defining the kind of syntactic annotations that will be used in the target corpus. In particular, they have contributed to defining a manual for the lowest annotation level (macro-syntax, disfluencies and paradigmatic grouping) as well as a strategy for semi-automatic micro-syntactic annotation, based on the reuse of an existing parser for written French. Kim Gerdes also has developped a few tools to experiment with syntactic annotation tasks.

7.2. Associate research team

Participant: Gérard Huet [correspondent].

Signes and an Indian group in computational linguistics lead by Amba Kulkarni (University of Hyderabad) and Puspak Battacharyya (IIT Mumbai) has been approved as an INRIA Franco-Indian Research Network in Computational Linguistics. This network is aimed at enforcing the relationship and exchanges in computational linguistics between France and India. In particular, a Sanskrit WordNet should be designed. See http://yquem.inria.fr/~huet/EA/Sanskrit08.html for details.

8. Dissemination

8.1. Involvement within the scientific community

8.1.1. Honours

- Gérard Huet has been a member of the *Académie des sciences* since November 2002.
- Gérard Huet has been a member of the *Academia Europae* since November 2002.

8.1.2. Editorial boards

- Gérard Huet is co-editing with Amba Kulkarni and Peter Scharf the proceedings of the 1st and 2nd International Sanskrit Computational Linguistics Symposiums, held respectively in INRIA Rocquencourt (France) in October 2007 and Brown University (Providence, Rhode Island, USA) in May 2008 [38]. Gérard Huet and Amba Kulkarni will also co-edit the proceedings of the 3rd Symposium, that will take place in University of Hyderabad (India) in January 2009 [37].
- Alain Lecomte has been on the editorial board of the international journal *Traitement Automatique des Langues (TAL)* since August 2001.
- Henri Portine is on the editorial board of the journal *ALSIC – Apprentissage des Langues et Systèmes d'Information et de Communication*.
- Christian Retoré has been a reviewer for *Mathematical Reviews* since October 2003.
- Christian Retoré has been editor in chief of the international journal *Traitement Automatique des Langues (TAL)* since April 2004 (and in the editorial board since 2001).
- Christian Retoré has been the editor in charge of the relations between mathematics, logic and computer science in La Gazette des Mathématiciens (quarterly, Société Mathématique de France) since July 2008.

8.1.3. Program committees of conferences and schools

- Maxime Amblard and Alain Lecomte were on the program committee of the Student Session of the European Summer School in Logic and Language and Information (ESSLLI), Hamburg, Germany, August 2008.
- Christian Bassac has been a member of the program committee of the 6th International Morphology Conference (Décembrettes-6, Bordeaux, France, December 2008).

- Gérard Huet was in the program committee of the 2nd Sanskrit Computational Linguistics Symposium, held at Brown University (Providence, Rhode Island, USA) in May 2008. He will co-chair with Amba Kulkarni the 3rd Symposium, that will take place in University of Hyderabad (India) in January 2009.

- Anna Kupść was a member of the program committee of the 15th International Conference on Head-Driven Phrase Structure Grammar (HPSG-2008, Keihanna, Japan, July 28–30, 2008), as well of the 16th International Conference on Intelligent Information Systems (IIS-2008, Zakopane, Poland, June 16–18, 2008).

- Alain Lecomte chaired with Myriam Quatrini the Workshop on Symmetric calculi and Ludics for the semantic interpretation at of ESSLLI-2008, Hamburg, Germany, August 2008. Richard Moot, Christian Retoré and Sylvain Salvati were on the program committee.

- Sylvain Salvati was a member of the program committee of the 13th conference on Formal Grammar (FG-2008, Hamburg, Germany, August 9–10, 2008).

- Christian Retoré chaired (with Laurent Prévot) the programme committee of the school in formal semantics (Atelier Jeunes Chercheurs en Sémantique et Modélisation) held in Toulouse in March 2008.

- Christian Retoré was a member of the program committee of IJCNLP 2009, The Third International Joint Conference on Natural Language Processing January 7-12, 2008, Hyderabad.

8.1.4. Academic committees

- Kim Guerdes was a member of the hiring committee for research and teaching assistants (maîtres de conférence) of the University Sorbonne nouvelle, 2008.

- Gérard Huet is a member of the Scientific Committee of the GIS SARIMA.

- Renaud Marlet was a member of the hiring committee for associate scientists (chargés de recherche) at INRIA Bordeaux - Sud-Ouest, 2008.

- Renaud Marlet was a member of the hiring committee for research and teaching assistants (maîtres assistants) at the École des Mines de Nantes, 2008.

- Since November 2007, Renaud Marlet has been a member of the Incentive Action Working Group (GTAI) of the Scientific and Technological Orientation Council (COST) of INRIA. Tasks include selecting and evaluating Collaborative Research Initiatives (ARC), Exploratory Actions (AE), Software Development Operations (ODL) and Technological Development Actions (ADT).

- Since November 2007, Renaud Marlet has been a member of the International Affairs Working Group (GTRI) of the Scientific and Technological Orientation Council (COST) of INRIA. Tasks include selecting and evaluating Associated Teams (EA) as well as proposals made in internatinal programs such as CONICyT/INRIA (with Chile), EuroMéditerrané 3+3 (with Algeria, Spain, Italy, Morocco, Tunisia), SECyT/INRIA-CNRS (with Argentina), STIC-AmSud (with South America), STIC-Tunisie (with Tunisia), and ERCIM post-doctorates.

- Since January 2008, Renaud Marlet has been a member of the teacher-and-researcher committee of INRIA Bordeaux - Sud-Ouest. Tasks include the evaluation of proposals concerning invitations of professors and temporary assignments at INRIA.

- Christian Retoré is a member of the hiring committee in computer science of Université Bordeaux 1.

- Christian Retoré has been a member of the committee of the faculty of mathematics and computer science (conseil d'UFR) of the Université Bordeaux 1 until May 2008.

- Christian Retoré is head of the Master on Algorithms and Formal Methods of Université Bordeaux 1.

- Christian Retoré is a member of the scientific committee (conseil scientifique) of the LaBRI.

- Christian Retoré was on the AERES evaluating committee of the ATILF lab (CNRS and Université de Nancy 2) which met in January 2008.

- Christian Retoré was on the AERES evaluating committee of the Laboratoire d'Informatique de Paris-Nord (CNRS and Université Paris 13) which met in January 2008.
- Christian Retoré was on the AERES evaluating committee of UMR-I Alpage (INRIA and Université Paris 7) which met in April 2008.
- Christian Retoré was on the AERES final committee for the computer science labs of Paris and the East of France.
- Christian Retoré has been a member of the Standing Committee of the European Summer School in Logic and Language and Information (ESSLLI) since August 2007, and left in August 2008 (not compatible with being the main organiser of ESSLLI 2009).

8.1.5. *Organization of events*

- Kim Gerdes organized a "Conférence du Samedi matin" at the Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA, Univ. Paris 3) in June 2008 on "les méthodes empiriques en syntaxe et l'émergence de nouveaux fonctionnalismes", with invited speaker Rens Bod from the University of Amsterdam (from exemplar to grammar: simulating language acquisition in U-DOP), followed by a discussion with Martin Kay, from Stanford University.
- Alain Lecomte organized a Prelude workshop in Autrans (France) in May 2008: Around Ludics, Dialogue, Game Theory and the question of questions.
- Alain Lecomte co-organized with Myriam Quatrini the Workshop on Symmetric calculi and Ludics for the semantic interpretation at of ESSLLI-2008, Hamburg, Germany, August 2008.
- The Signes projet-team is the main organizer of the European Summer School in Logic and Language and Information (ESSLLI), Bordeaux, France, July 2009.
- Christian Retoré organized (with Laurent Prévot) the the school in formal semantics (Atelier Jeunes Chercheurs en Sémantique et Modélisation) held in Toulouse in March 2008.
- Maxime Amblard is member of the organization committee of the event "50 ans de l'Association pour le Traitement Automatique des Langues (ATALA)", Paris, France, June 2009.

8.1.6. *Miscellaneous*

- Maxime Amblard is vice-treasurer of the Association pour le Traitement Automatique des Langues (ATALA).

8.2. Teaching

As many of its members are university staff, Signes is intensively implied in teaching, both in the computer science cursus (University Bordeaux 1) and in the linguistic cursus (University of Bordeaux 3). Signes is also teaching in summer schools for PhD students and colleagues. What follows only covers lectures whose topic is computational linguistics:

- Lionel Cément taught linguistic formalization to linguistics students at the University Bordeaux 3. The aim was to present the Lexical Functional Grammar (LFG) theory as well as an implementation of a few linguistic phenomena with the XLFG5 parser (see section 5.5).
- Lionel Clément and Christian Retoré are teaching a lecture on Natural Language Processing (48h, 4th year), Université Bordeaux 1.
- Gérard Huet gave a course on algebraic and functional calculi at the Master Parisien de Recherche en Informatique (MPRI, Master 2, 12h).
- Anna Kupść is responsible for the Master programme in *Computational Linguistics* at University of Bordeaux 3. She taught two one-semester Master courses (*Linguistics and Programming Languages* and *Linguistics and Formalisation*), a two-semester course (Licence 3, *Computational Linguistics*), and a semester course (Licence 2, *Introduction to Computational Lingustics*).
- Christian Retoré is teaching (with Bruno Courcelle) a course on Logic and languages (1/2 $times$ 24h, 5th year), Université Bordeaux 1.

8.3. Defended Theses

- Émilie Voisin defended her PhD thesis at University Bordeaux 3 on November 27th, 2008: *analyse syntaxique et formalisation d'énoncés en Langue des Signes Française* (syntactic analysis and formalization of utterances in French Sign Language).

8.4. Thesis Committees

- Mauro Gaio was in the PhD committee (supervisor with Thierry Nodenot) of Pierre Loustau, who defended his thesis at the Université de Pau et des Pays de l'Adour on November 14th, 2008: *from displacements to itinerary, from clauses to discourse.*

- Alain Lecomte (rapporteur) and Christian Retoré (président) were in the PhD committee (examinateur) of Franck Sablé, who defended his thesis (dir. Francis Corblin) at Université Paris 4 on December 5th: *Sémantique suppositionnelle et différentielle de l'algèbre discursive d(S) appliquée aux connecteurs* et, mais, si, donc.

- Renaud Marlet was in the PhD committee (examinateur) of Nicolas Palix, who defended his thesis at University Bordeaux 1 on September 17th, 2008: *langages dédiés au développement de services de communications* (domain-specific languages for the developement of communication services).

- Henri Portine and Renaud Marlet were in the PhD committee (co-supervisors) of Émilie Voisin, who defended her thesis at University Bordeaux 3 on November 27th, 2008: *analyse syntaxique et formalisation d'énoncés en Langue des Signes Française* (syntactic analysis and formalization of utterances in French Sign Language).

- Christian Retoré was on the PhD committee (rapporteur) of Sébastien Hinderer, who defended his thesis at Université Nancy 1 on October 21st: *automatisation de la construction sémantique dans TY*n.

- Christian Retoré was on the PhD committee (rapporteur) of Céline Raynal, who defended her thesis at Université Paris 7 on June 27th: *la restriction en français, trois études sémantiques*

8.5. Academic supervision

8.5.1. Student internship supervision – fourth and fifth year

- Christian Retoré and Sylvain Salvati supervised the Master thesis of Arthur Ball (ENSIMAG) on *syntactic analysis and Datalog queries.*

8.5.2. PhD supervision

- Christian Bassac is supervising the PhD thesis work of Olivier Taïs (Université Bordeaux 3): *the dynamics of the lexicon.*

- Christian Bassac and Christian Retoré are co-supervising the PhD thesis work of Bruno Mery (Université Bordeaux 1, ministry grant): *type theory for lexical semantics.*

- Mauro Gaio supervised with Thierry Nodenot the PhD thesis work of Pierre Loustau (Université de Pau et des Pays de l'Adour): *from displacements to itinerary, from clauses to discourse.*

- Mauro Gaio is supervising with Christian Sallaberry the PhD thesis work of Nguyen Van Tien (Université de Pau et des Pays de l'Adour): *improving a geographical ontology: a method based on a semantic corpus analysis.*

- Alain Lecomte is supervising the PhD thesis of Christophe Onambele (Université Paris 8): *Minimalist Grammars and application to linguistic descriptions.*

- Alain Lecomte is co-supervising the PhD thesis of Mawusse Kpakpo Akue Adotevi (Université de Lomé, Togo): *Dialogical Games and Language Games*

- Henri Portine and Renaud Marlet co-supervised the PhD thesis of Émilie Voisin (University Bordeaux 3): *syntactic analysis and formalization of utterances in French Sign Language*.

- Christian Retoré and Sylvain Salvati are co-supervising the PhD thesis of Pierre Bourreau (Université Bordeaux 1, INRIA CORDI grant): *the treatment and use of non-linearity in computationnal linguistics*.

8.6. Participation to colloquia, seminars, invitations

8.6.1. Visiting scientists

- Prof. Denis Delfitto (Università degli studi di Verona, linguistics, generative syntax and formal semantics) was invited for two months in the Signes group.

8.6.2. Talks at conferences, seminar talks and invitations

- Maxime Amblard presented his work at Calligramme project-team seminar in Nancy in May 2008: Interface syntaxe-sémantique dans un cadre logique.

- Pierre Boureau visited Makoto Kanazawa at the NII (Tokyo, Japan) in July 2008.

- Pierre Bourreau, Lionel Clément, Bruno Mery, Richard Moot, Christian Retoré and Sylvain Salvati gave talks to high school students for the Fête de la Science in November: Un ordinateur peut-il comprendre notre langue ?

- Kim Gerdes gave a talk the 9th Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008, Lyon) in March 2008: "L'alignement pour les pauvres : adapter la bonne métrique pour un algorithme dynamique de dilatation temporelle pour l'alignement sans ressources de corpus bilingues".

- Kim Gerdes gave a talk (with Pollet Samvelian) at the Conference on Complex Predicates in Iranian Languages (Paris) in July 2008: automatic extraction and partial annotation of persian complex predicates. He also gave a seminar talk on "Les constructions à verbe support du persan : une approche statistique".

- Kim Gerdes gave a talk at The International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008, Hangzhou, China) in September 2008: poverty driven bilingual alignment.

- Kim Gerdes gave a talk at the 27th Colloque International Lexique et Grammaire 2008 (L'Aquila, Italia) in September 2008: a statistical approach to persian light verb constructions.

- Kim Gerdes gave a talk (with Miao Jun) at the Colloque international Fu Lei et traduction (Nanjing, Chine) in May 2008: Donner accès à l'oeuvre de Fu Lei.

- Gérard Huet gave a talk at the NLP Winter School at IIIT Hyderabad (India) in Januray: An overview of the Zen computational linguistics toolkit.

- Gérard Huet gave a talk at the 2nd International Symposium on Sanskrit Computational Linguistics at Brown University (Providence, Rhode Island, USA) in May: Formal structure of Sanskrit text: requirements analysis for a mechanical Sanskrit processor.

- Gérard Huet gave a talk at the Exalead company (Paris) in June: Technologie Zen de machines d'état fini et applications au traitement de la langue naturelle.

- Gérard Huet gave a invited lecture at the Constraints and Language Processing workshop (CSLP-2008) at ESSLLI (Hamburg, Germany) in August: Syntax-Semantics Interface for Sanskrit Using Constraint Processing of Semantic Roles.

- Gérard Huet gave a talk at the Laboratoire d'Informatique de Nantes Atlantique in November: Machines d'Eilenberg, bibliothèque Zen, et applications au traitement informatique du sanskrit.

- Gérard Huet and Benoît Razet presented a tutorial at the 6th International Conference on Natural Language Processing (ICON-2008) in Pune (India) in December: Eilenberg machines, the Zen toolkit, and applications to Sanskrit Computational Linguistics.

- Anna Kupść gave a talk at the TALC Seminar (LORIA, Nancy, France) in October: Adjectives in TreeLex.

- Anna Kupść gave a talk (with Cécile Fabre) at the Master TAL Seminar (ERSS, Toulouse, France) in November: Extraction de cadres de sous-catégorisation pour les adjectifs.

- Alain Lecomte was invited to give a talk at the University Polytecnica de Barcelona at a workshop organized by Glyn Morrill in December 2007 (which was not mentioned in the 2007 report).

- Alain Lecomte, Richard Moot and Sylvain Salvati were invited to (each) give a talk at the workshop "50 years of the syntactic calculus" organized by Claudia Casadia at the University of Chieti (Italy) to celebrate the 50th anniversary of the seminal article by J. Lambek *On the calculus of syntactic types*. Sylvain Salvati presented his work on derivations of Lambek grammars, and Richard Moot on Lambek grammars and hyperedge replacement grammars.

- Alain Lecomte was invited to give a talk during the "Journées Logique de l'Interaction et Géométrie de la Cognition (LIGC)" organized by J.-Y. Girard's group in Carry-le-Rouet (France) in October.

- Renaud Marlet presented a poster at the 15th Conférence sur le Traitement Automatique des Langues Naturelles (TALN '08, Avignon, France) in June: Un sens logique pour les graphes sémantiques (a logical meaning for semantic graphs).

- Richard Moot gave a talk at the The 9th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+9) in Tübingen (Germany) in June: Lambek grammars, tree adjoining grammars and hyperedge replacement grammars.

- Richard Moot gave an invited talk "Proof nets for the Lambek-Grishin Calculus" at a meeting of the ANR PRELUDE project in Paris, in January.

- Richard Moot gave an invited talk "Type-Logical and Hyperedge Replacement Grammars" at the MF (Méthodes Formelles) group seminar (LaBRI, Bordeaux, France) in May.

- Daniele Porello presented his work on Generative Lexicon and Reasoning Frames at AISC Congress (Associazione Italiana Scienze Cognitive), Turin, December 2008.

- Daniele Porello presented his work on *Preferences Update and Deliberation about Dimensions*, SIFA 08 (Italian Society for Analytic Philosophy), Bergamo, September 2008 (Accepted).

- Daniele Porello was accepted to present his work on *Deliberation and Changing Dimensions* at COLLOC (Change of Logic, Logic of Change) Praha, September 2008. (Accepted).

- Daniele Porello presented his work on *A translation of Discursive Dilemma into Condorcet's Paradox* at ECAP 08 (European Conference for Analytic Philosophy) Krakow, August, 2008.

- Daniele Porello presented his work on *Lexical semantics, Generative Lexicon and Reasoning Frames* at GESCO 08 (GEnova Scienze COgnitive), Genova 2008.

- Daniele Porello presented his joint work with Christian Retoré on *Proof nets without links* at *50 Yers of Syntactic Calculus*, Chieti, July 2008

- Benoît Razet gave a talk at the 2nd workshop on Mathematically Structured Functional Programming (MSFP) in Reykjavik (Iceland) in July: Simulating finite Eilenberg machines with a reactive engine.

- Benoît Razet gave a talk at the 13th International Conference on Implementation and Applications of Automata (CIAA) in San Francisco (CA, USA) in July: Finite Eilenberg machines.

- Benoît Razet presented his work at the LORIA (Nancy, France), at LaBRI (Bordeaux, France) in October, and at the Microsoft Research-INRIA Joint Centre (Orsay, France) in December: Modélisation et simulation des automates à l'aide de machines d'Eilenberg.

- Christian Retoré gave a talk at the Groupe de Travail Graphes et Applications (LaBRI) on January: Proofnets, cographs and perfect matchings.

- Christian Retoré gave a lecture at the École d'Automne de Linguistique (Ecole Normale Supérieure) in September: Categorial grammars for computing the correspondence between syntax and semantics — an example of interplay between mathematical logic and computational linguistics.

- Christian Retoré gave a talk at a workshop on Relations (Maison des Sciences de l'Homme de Paris Nord) in December: Les relations dans les modèles de la logique linéaire.

- Sylvain Salvati was invited to present his work on linear logic, automata and minimalist grammars at the IML seminar (Marseille, France) in January.

- Sylvain Salvati presented his work on the derivations of Lambek grammars and their relation with hypergraph grammars and monadic second order logic at the MF (Méthodes Formelles) group seminar (LaBRI, Bordeaux, France) in January.

- Sylvain Salvati presented his work on recognizable sets of lambda-terms at the MF (Méthodes Formelles) group seminar (LaBRI, Bordeaux, France) in February.

- Sylvain Salvati was invited to present his work concerning the relation between minimalist grammars and logic (linear logic and monadic second order logic) at the 5th workshop on lambda calculus and formal grammars (Tokyo, Japan) in March.

- Sylvain Salvati was invited to present his work on the derivations of Lambek grammars for the 50th anniversary of Lambek's original paper on the syntactic calculus (Chieti, Italy) in July.

- Sylvain Salvati was invited to present his work on recognizable sets of lambda-terms and their closure properties at the 6th workshop on lambda calculus and formal grammars (Haifa, Israel) in October.

- Émilie Voisin and Maxime Amblard gave a talk at the workshop on the Traitement Automatique des Langues des Signes (TALS, Avignon, France) in June: Modélisation de la LSF en vue d'une automatisation.

- Émilie Voisin gave an invited talk about her work on french sign language at a seminar of the LPL team (University of Aix-en-Provence, France) in April.

- Émilie Voisin was invited to present her work at a workshop on typology and grammatical roles held at University of Aix-en-Provence in November.

8.6.3. Participation to conferences and summer schools

- Maxime Amblard, Renaud Marlet and Christian Retoré attended the 15th Conférence sur le Traitement Automatique des Langues Naturelles (TALN '08), Avignon, France, June 2008.

- Maxime Amblard, Émilie Voisin and Christian Retoré attended the workshop on the Traitement Automatique des Langues des Signes (TALS '08), Avignon, France, June 2008.

- Pierre Bourreau, Denis Delfitto and Renaud Marlet, Daniele Porello and Christian Retoré attended the Journées de Sémantique et Modélisation (JSM '08), Toulouse, France, April 2008.

- Pierre Bourreau, Bruno Mery, Daniele Porello and Christian Retoré attended the 20th European Summer School in Logic, Language and Information (ESSLLI-2008), Hamburg, Germany, August 2008.

- Alain Lecomte participated to the Workshop on Symmetric calculi and Ludics for the semantic interpretation at of ESSLLI-2008, Hamburg, Germany, August 2008.

- Alain Lecomte participated to the workshop "Propositions: Ontology, Philosophy, Pragmatics" organized by Frederike Moltmann in Venice, Italy, November 2008.

- Alain Lecomte, Richard Moot and Sylvain Salvati participated to the workshop "50 years of the syntactic calculus" organized by Claudia Casadia at the University of Chieti (Italy) to celebrate the 50th anniversary of the seminal article by J. Lambek *On the calculus of syntactic types*.

9. Bibliography

Major publications by the team in recent years

[1] C. BASSAC. *Principes de morphologie anglaise*, Linguistica, Presses Universitaires de Bordeaux, 2004.

[2] C. BASSAC. *Morphologie et Information Lexicale*, Mémoire d'habilitation à diriger des recherches, Université Michel de Montaigne Bordeaux 3, December 2006.

[3] L. CLÉMENT, A. KINYON. *Generating parallel multilingual LFG-TAG grammars from a MetaGrammar*, in "Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Sapporo, Japan", July 2003.

[4] K. GERDES. *Topologie et grammaires formelles de l'allemand*, Thèse de Doctorat, Université Paris 7, 2002.

[5] G. HUET. *Transducers as lexicon morphisms, phonemic segmentation by euphony analysis, application to a sanskrit tagger*, in "Journal of Functional Programming", vol. 15, n° 4, July 2005, p. 573–614.

[6] A. LECOMTE. *Categorial Grammar for Minimalism*, in "Logic and Grammar", C. CASADIO, P. SCOTT, R. SEELY (editors), CSLI Pulications, 2005.

[7] J. LESBEGUERIES, M. GAIO, P. LOUSTAU. *Geographical information access for non-structured data*, in "Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France, April 23-27, 2006", H. HADDAD (editor), ACM, 2006, p. 83–89, http://doi.acm.org/10.1145/1141277.1141296.

[8] R. MOOT. *Proof nets for linguistic analysis*, Ph. D. Thesis, UIL-OTS, Universiteit Utrecht, 2002.

[9] R. MOOT. *Automated Extraction of Type-Logical Supertags from the Spoken Dutch Corpus*, in "The Complexity of Lexical Descriptions and its Relevance to Natural Language Processing: A Supertagging Approach", S. BANGALORE, A. JOSHI (editors), MIT Press, 2007.

[10] E. NYBERG, T. MITAMURA, J. P. CALLAN, J. G. CARBONELL, R. E. FREDERKING, K. COLLINS-THOMPSON, L. HIYAKUMOTO, Y. HUANG, C. HUTTENHOWER, S. JUDY, J. KO, A. KUPSC, L. V. LITA, V. PEDRO, D. SVOBODA, B. V. DURME. *The JAVELIN Question-Answering System at TREC 2003: A Multi-Strategh Approach with Dynamic Planning*, in "TREC", 2003.

[11] H. PORTINE. *La syntaxe de Damourette et Pichon comme outil de représentation du sens*, in "Modèles linguistiques", vol. 23, n° 2, 2002, p. 21–46.

[12] C. RETORÉ. *Logique linéaire et syntaxe des langues*, Mémoire d'habilitation à diriger des recherches, Université de Nantes, January 2002.

[13] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", vol. 2, n° 1, 2004, p. 3–25.

[14] S. SALVATI. *Problèmes de filtrage et problèmes d'analyse pour les grammaires catégorielles abstraites*, Ph. D. Thesis, Institut National Polytechnique de Lorraine, 2005.

Year Publications

Articles in International Peer-Reviewed Journal

[15] C. BASSAC, M. CIÇEK. *Factorisation des suffixes de pluriel et de possessif en turc*, in "Revue de Sémantique et Pragmatique", 2009.

[16] C. BASSAC, B. MERY, C. RETORÉ. *Towards a type-theoretical account of lexical semantics*, in "Journal of Logic, Language and Information", To appear, 2009.

[17] P. LOUSTAU, T. NODENOT, M. GAIO. *Design principles and first educational experiments of πR, a platform to infer geo-referenced itineraries from travel stories*, in "Journal of Interactive Technology and Smart Education (ITSE)", To appear, 2008.

[18] C. RETORÉ. *Note de lecture sur "Two-Step Approaches to Natural Language Formalisms" (Frank Morawietz, 2003)*, in "Traitement Automatique des Langues (TAL)", To appear, vol. 48, n⁰ 3, 2008.

[19] C. RETORÉ, S. SALVATI. *A Faithful Representation of Non-Associative Lambek Grammars in Abstract Categorial Grammars*, in "Journal of Logic, Language and Information", To appear, 2009.

[20] S. SALVATI. *On the membership problem for non-linear Abstract Categorial Grammars*, in "Journal of Logic, Language and Information", To appear, 2009.

Articles in National Peer-Reviewed Journal

[21] P. LOUSTAU, M. GAIO, T. NODENOT. *Interprétation automatique d'itinéraires à partir d'un corpus de récits de voyages pilotée par un usage pédagogique*, in "Revue des Nouvelles Technologies de l'Information", vol. E-13, 2008, p. 177–206.

[22] C. RETORÉ. *Les mathématiques de la linguistique computationnelle. Deuxième volet: Logique*, in "La Gazette des mathématiciens", vol. 116, April 2008, p. 29–63.

[23] C. RETORÉ. *Les mathématiques de la linguistique computationnelle. Premier volet: la théorie des langages*, in "La Gazette des mathématiciens", vol. 115, January 2008, p. 35–62.

International Peer-Reviewed Conference/Proceedings

[24] M. AMBLARD, J. HEINECKE, E. MAILLEBUAU. *Discourse Representation Theory et graphes sémantiques : formalisation sémantique en contexte industriel*, in "15e Conférence sur le Traitement Automatique des Langues Naturelles (TALN'08), Avignon, France", Poster, June 2008, p. 350–359.

[25] K. GERDES. *L'alignement pour les pauvres : adapter la bonne métrique pour un algorithme dynamique de dilatation temporelle pour l'alignement sans ressources de corpus bilingues*, in "9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008), Lyon, France", S. HEIDEN, B. PINCEMIN (editors), Presses Universitaires de Lyon (PUL), March 2008.

[26] K. GERDES. *Poverty Driven Bilingual Alignment*, in "International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008), Hangzhou, China", R. XIAO, L. HE, M. YUE (editors), September 2008.

[27] K. GERDES, P. SAMVELIAN. *A Statistical Approach to Persian Light Verb Constructions*, in "Actes du 27e Colloque International sur le lexique et la grammaire 2008, L'Aquila, Italy", To appear, September 2008.

[28] G. HUET. *Formal structure of Sanskrit text: Requirements analysis for a mechanical Sanskrit processor*, in "Topics in Sanskrit Computational Linguistics", G. HUET, A. KULKARNI, P. SCHARF (editors), Lecture Notes in Computer Science (LNCS), To appear, Springer, 2009.

[29] A. LECOMTE. *Semantics in Minimalist-Categorial Grammar*, in "Proceedings of the 13th conference on Formal Grammar, Hamburg, Germany", P. DE GROOTE, L. KALLMEYER, G. PENN, G. SATTA (editors), CSLI Publications, August 2008.

[30] P. LOUSTAU, T. NODENOT, M. GAIO. *Spatial decision support in the pedagogical area: Processing travel stories to discover itineraries hidden beneath the surface*, in "11th AGILE International Conference on Geographic Information Science, Girona, Spain", Lecture Notes in Geoinformation and Cartography, Springer-Verlag, May 2008.

[31] R. MARLET. *Un sens logique pour les graphes sémantiques*, in "15e Conférence sur le Traitement Automatique des Langues Naturelles (TALN'08), Avignon, France", Poster, June 2008, p. 498–507.

[32] R. MOOT. *Lambek grammars, tree adjoining grammars and hyperedge replacement grammars*, in "9th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+9), Tübingen, Germany", C. GARDENT, A. SARKAR (editors), June 2008, p. 65–72.

[33] B. RAZET. *Finite Eilenberg Machines*, in "Proceedings of the 13th International Conference on Implementation and Application of Automata (CIAA 2008), San Francisco, USA", O. H. IBARRA, B. RAVIKUMAR (editors), Lecture Notes in Computer Science (LNCS), n⁰ 5148, Springer-Verlag, July 2008, p. 242–251.

[34] B. RAZET. *Simulating Finite Eilenberg Machines with a Reactive Engine*, in "Proceedings of the workshop on Mathematically Structured Functional Programming (MSFP 2008), Reykjavik, Iceland", Electric Notes in Theoretical Computer Science, July 2008.

[35] C. SALLABERRY, M. GAIO, D. PALACIO, J. LESBEGUERIES. *Fuzzying GIS Topological Functions for Geographical Information Retrieval Needs*, in "2nd international workshop on Geographical Information Retreival (GIR'08), satellite of the 17th ACM Conference on Information Knowledge Management (CIKM 2008), Napa Valley, California, USA", ACM Press, 2008, p. 1–8.

Workshops without Proceedings

[36] N. ASHER, P. MULLER, M. GAIO. *Spatial entities are temporal entities too: the case of motion verbs*, in "Workshop on Methodologies and Resources for Processing Spatial Language, satellite of the 6th international conference on Language Resources and Evaluation (LREC'08), Marrakech, Maroc", may 2008, p. 16–21.

Books or Proceedings Editing

[37] G. HUET, A. KULKARNI (editors). *3rd Sanskrit Computational Linguistics Symposium*, Lecture Notes in Computer Science (LNCS), To appear, Springer, 2009.

[38] G. HUET, A. KULKARNI, P. SCHARF (editors). *Topics in Sanskrit computational linguistics*, Lecture Notes in Computer Science (LNCS), To appear, Springer, 2009.

Research Reports

[39] A. BALL. *Analyse syntaxique et requêtes Datalog*, Rapport de stage de Master, INP Grenoble-ENSIMAG et INRIA, 2008.

Other Publications

[40] G. HUET, B. RAZET. *Computing with Relational Machines*, Tutorial given at the 6th International Conference on Natural Language Processing (ICON 2008), Pune, India, December 2008.

References in notes

[41] M. AMBLARD. *Calcul de représentations sémantiques et syntaxe générative : les Grammaires Minimalistes Catégorielles*, Ph. D. Thesis, Université de Bordeaux 1, September 2007.

[42] M. AMBLARD, É. VOISIN. *Vers une analyse automatique de la LSF : une modélisation de la Langue des Signes Française par les Grammaires Minimalistes*, Technical report, n[o] HAL INRIA-00186965, INRIA, 2007.

[43] M. C. BAKER. *Thematic roles and syntactic structure*, in "Elements of grammar: Handbook of generative syntax, Dordrecht", L. HAEGEMAN (editor), Kluwer, 1997, p. 73–137.

[44] K. R. BEESLEY, L. KARTTUNEN. *Finite-State Morphology: Xerox Tools and Techniques*, Cambridge University Press, 2002.

[45] P. BLACHE. *Property Grammars: A Fully Constraint-based Theory*, in "Constraint Satisfaction and Language Processing", H. CHRISTIANSEN, P. SKADHAUGE, J. VILLADSEN (editors), Springer, 2005.

[46] M.-H. CANDITO. *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées, Application au français et à l'italien.*, Ph. D. Thesis, Université de Paris 7, 1999.

[47] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *A new metagrammar compiler*, in "TAG+6", 2002.

[48] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Représentation et gestion du lexique d'une grammaire d'arbres adjoints*, in "Traitement automatique des langues", vol. 43, n° 3, 2004.

[49] A. DIKOVSKY, L. MODINA. *Dependencies on the other side of the Curtain*, in "Traitement Automatique des Langues", vol. 41, n° 1, 2000, p. 67-95.

[50] M.-L. GUÉNOT, P. BLACHE. *A descriptive and formal perspective for grammar development*, in "Workshop on Foundations of Natural-Language Grammars, Edimbourg, Royaume-Uni", 2005.

[51] G. HUET. *Linear Contexts and the Sharing Functor: Techniques for Symbolic Computation.*, in "Thirty Five Years of Automating Mathematics", F. KAMAREDDINE (editor), Kluwer, 2003.

[52] G. HUET. *Design of a Computational Linguistics Platform for Sanskrit*, in "Logical Aspects of Computational Linguistics, LACL 05", Invited Lecture, 2005.

[53] G. HUET. *Un système de traitement informatique du sanskrit*, in "Journée ATALA : Traitement Automatique des Langues Anciennes", Invited Lecture, 2005.

[54] G. HUET. *Shallow syntax analysis in Sanskrit guided by semantic nets constraints*, in "International Workshop on Research Issues in Digital Libraries (IWRIDL 2006), Kolkata, India", To appear in ACM Digital Library, December 2007.

[55] R. JACKENDOFF. *Languages of the mind*, MIT Press, 1995.

[56] A. JOSHI, L. LEVY, M. TAKAHASHI. *Tree Adjunct Grammar*, in "Journal of Computer and System Sciences", vol. 10, 1975, p. 136–163.

[57] A. JOSHI, Y. SCHABES. *Tree Adjoining Grammars*, in "Handbook of Formal Languages, Berlin", G. ROZENBERG, A. SALOMAA (editors), vol. 3, chap. 2, Springer Verlag, 1996.

[58] A. JOSHI, K. VIJAY-SHANKER, D. WEIR. *The convergence of mildly context-sensitive grammar formalisms*, in "Fundational issues in natural language processing", P. SELLS, S. SCHIEBER, T. WASOW (editors), MIT Press, 1991.

[59] M. JUN, K. GERDES. *Donner accès à l'oeuvre de Fu Lei*, in "Actes du Colloque international Fu Lei et traduction, Nanjing, China", X. JUN (editor), May 2008.

[60] H. KAMP, U. REYLE. *From Discourse to Logic*, D. Reidel, Dordrecht, 1993.

[61] M. KANAZAWA. *A prefix-correct Earley recognizer for multiple context-free grammars*, in "9th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+9), Tübingen, Germany", C. GARDENT, A. SARKAR (editors), June 2008, p. 49–56.

[62] R. M. KAPLAN, M. KAY. *Regular Models of Phonological Rule Systems*, in "Computational Linguistics", vol. 20,3, 1994, p. 331–378.

[63] L. KARTTUNEN. *Applications of Finite-State Transducers in Natural Language Processing*, in "Proceedings, CIAA-2000", 2000.

[64] K. KOSKENNIEMI. *A general computational model for word-form recognition and production*, in "10th International Conference on Computational Linguistics", 1984.

[65] A. KRATZER. *On external arguments*, in "Functional Projections, Amherst, USA", E. BENEDICTO, J. RUNNER (editors), University of Massachussets, Occasional Papers, 1994, p. 103–130.

[66] A. LECOMTE, C. RETORÉ. *Extending Lambek grammars: a logical account of minimalist grammars*, in "Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, ACL 2001, Toulouse", ACL, July 2001, p. 354–361.

[67] A. LECOMTE, C. RETORÉ. *Towards a Minimal Logic for Minimalist Grammars: a Transformational Use of Lambek Calculus*, in "Formal Grammar (FG 1999)", FoLLI, 1999, p. 83–92.

[68] D. LILLO-MARTIN, E. S. KLIMA. *Pointing out differences: ASL pronouns in syntactic theory*, in "Theoretical issues in sign language research – Vol 1 Linguistics", S. D. FISHER, P. SIPLE (editors), University of Chicago Press, 1990, p. 191–210.

[69] D. LILLO-MARTIN. *Universal Grammar and American Sign Language: Setting the Null Argument Parameters*, Kluwer, 1991.

[70] C. MANNING, H. SCHUTZE. *Foundations of statistical natural language processing*, MIT Press, 1999.

[71] I. MELCUK. *Dependency syntax – theory and practice*, Linguistics, State University of New York Press, 1988.

[72] J. MICHAELIS. *Derivational minimalism is mildly context sensitive*, in "Logical Aspects of Computational Linguistics (LACL 1998), selected papers", M. MOORTGAT (editor), LNCS/LNAI, n$^{\text{o}}$ 2014, Springer-Verlag, 2001, p. 179–198.

[73] M. MOORTGAT. *Categorial Type Logic*, in "Handbook of Logic and Language, Amsterdam", J. VAN BENTHEM, A. TER MEULEN (editors), chap. 2, North-Holland Elsevier, 1996, p. 93–177.

[74] F. MORAWIETZ. *Two-Step Approaches ot Natural Language Formalisms*, Studies in Generative Grammar, Mouton de Gruyter, Berlin · New York, 2003.

[75] U. MÖNNICH, J. MICHAELIS, F. MORAWIETZ. *On Minimalist Attribute Grammars and Macro Tree Transducers*, in "Linguistic Form and its Computation", C. ROHRER, A. ROSSDEUTSCHER, H. KAMP (editors), CSLI Publications, 2004.

[76] C. NEIDLE, J. KEGL, D. MACLAUGHLIN, B. BAHAN, R. G. LEE. *The Syntax of American Sign Language – Functional Categories and Hierarchical Structure*, MIT Press, 2000.

[77] S. POGODALLA. *Computing Semantic Representation: Towards ACG Abstract Terms as Derivation Trees*, in "Proceedings of the Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+7)", May 2004, p. 64–71.

[78] G. K. PULLUM, B. C. SCHOLZ. *On the distinction bewteen moedl-theoretic and generative-enumerative syntax*, in "Logical Aspects of Computational Linguistics (LACL 2001)", P. DE GROOTE, G. MORRILL, C. RETORÉ (editors), LNCS/LNAI, n$^{\text{o}}$ 2099, Springer-Verlag, 2001, p. 17–43.

[79] J. PUSTEJOVSKY. *The Generative Lexicon*, MIT Press, 1995.

[80] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", vol. 2, n$^{\text{o}}$ 1, 2004, p. 3–25.

[81] E. STABLER. *Derivational Minimalism*, in "Logical Aspects of Computational Linguistics (LACL 1996)", C. RETORÉ (editor), LNCS/LNAI, vol. 1328, Springer-Verlag, 1997, p. 68–95.

[82] E. STABLER. *Remnant movement and structural complexity*, in "Constraints and Resources in Natural Language Syntax and Semantics", G. BOUMA, E. HINRICHS, G.-J. M. KRUIJFF, R. OEHRLE (editors), distributed by Cambridge University Press, CSLI, 1999, p. 299–326.

[83] V. Sutton. *Lessons in SignWriting*, Deaf Action Committee for SignWriting, La Jollla, CA, 2002, http://www.signwriting.org.

[84] J. van Eijck, H. Kamp. *Representing Discourse in Context*, in "Handbook of Logic and Language, Amsterdam", J. van Benthem, A. ter Meulen (editors), chap. 3, North-Holland Elsevier, 1996, p. 179–237.